

Figure 1

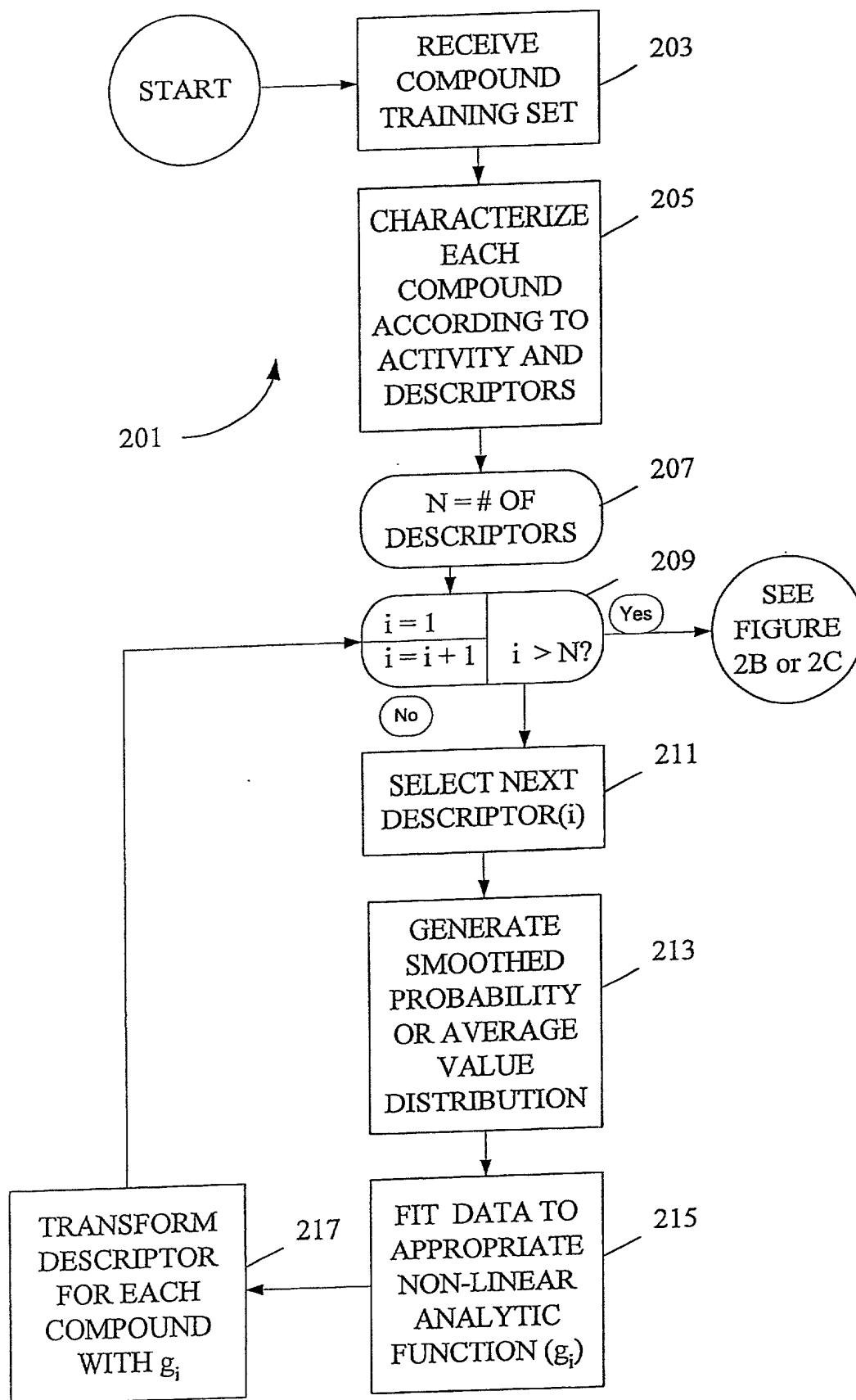


FIGURE 2A

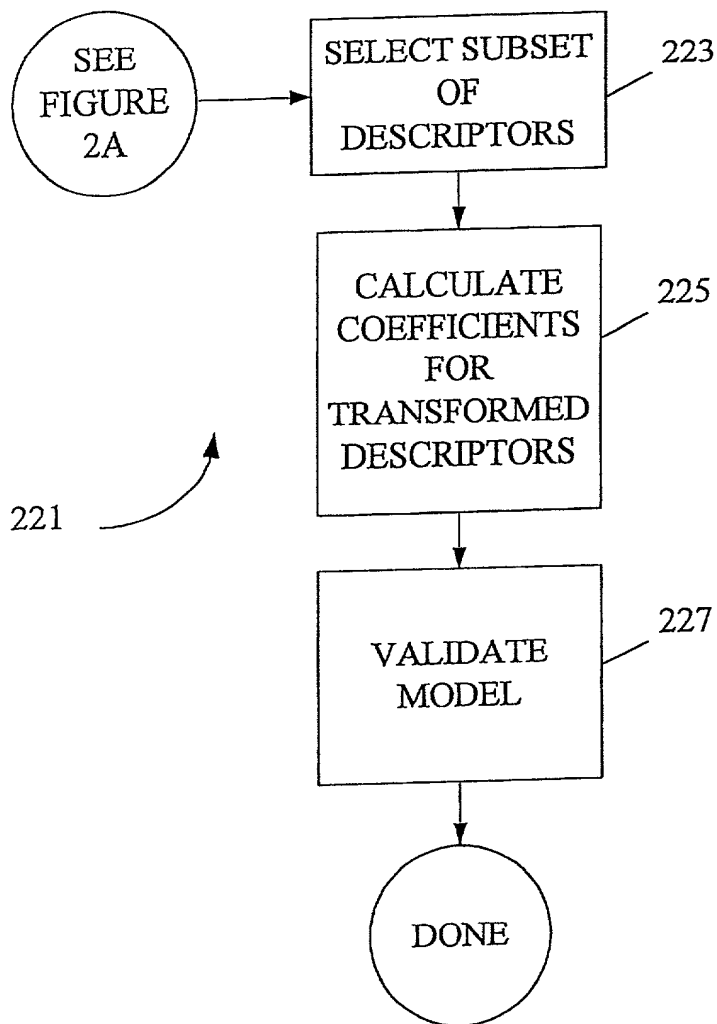


FIGURE 2B

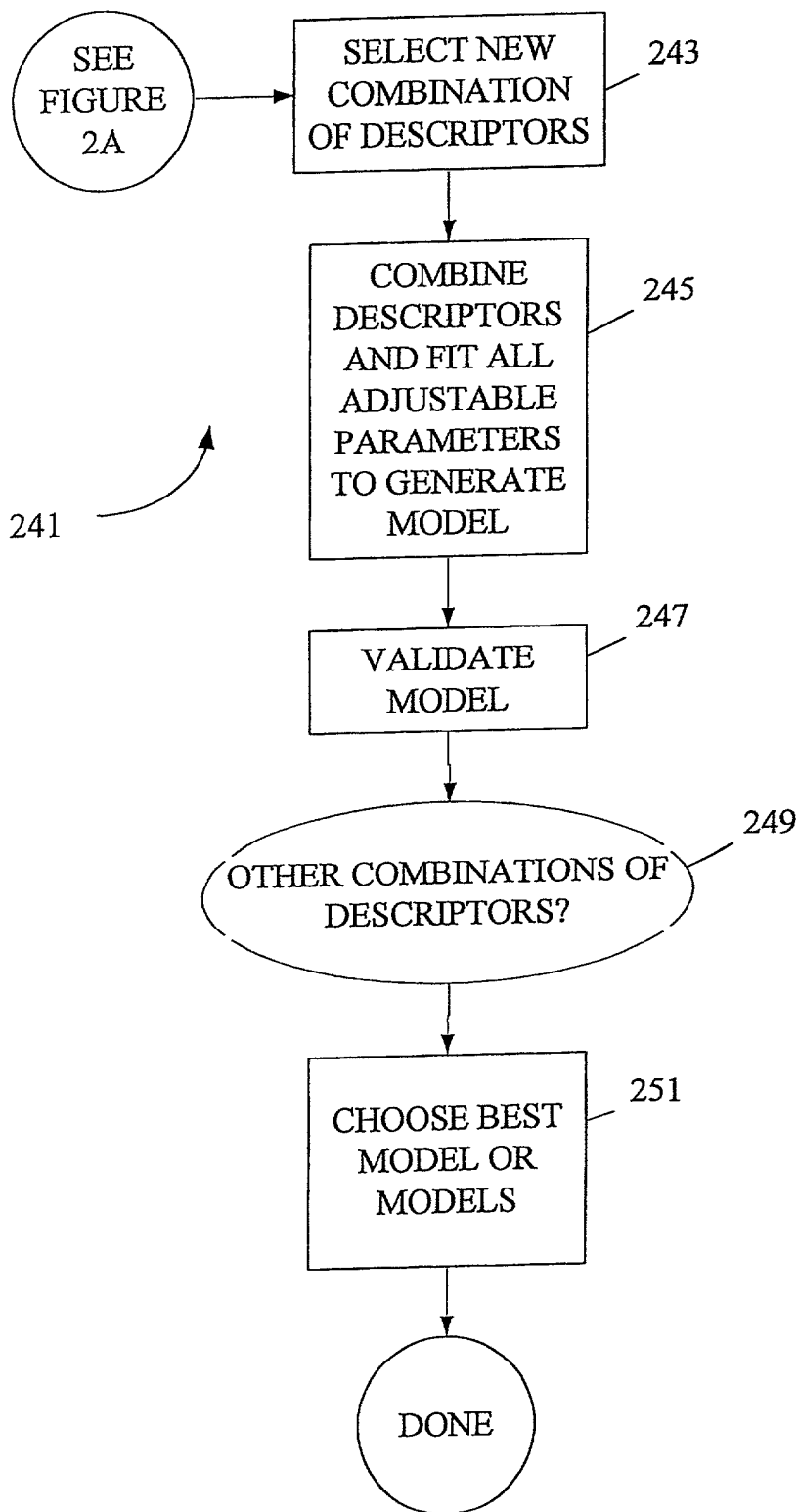
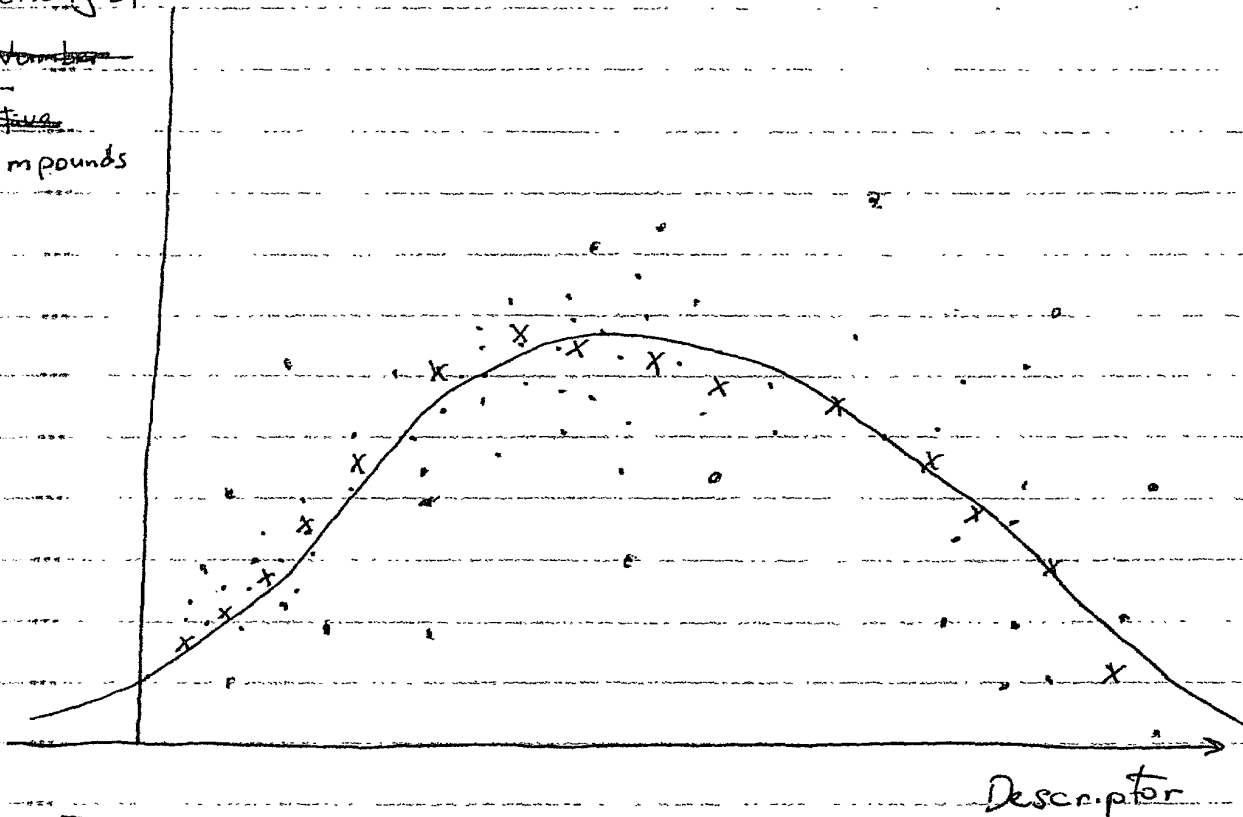


FIGURE 2C

100225003 1401901

Activity of
~~Number~~
~~of~~
~~Active~~
Compounds

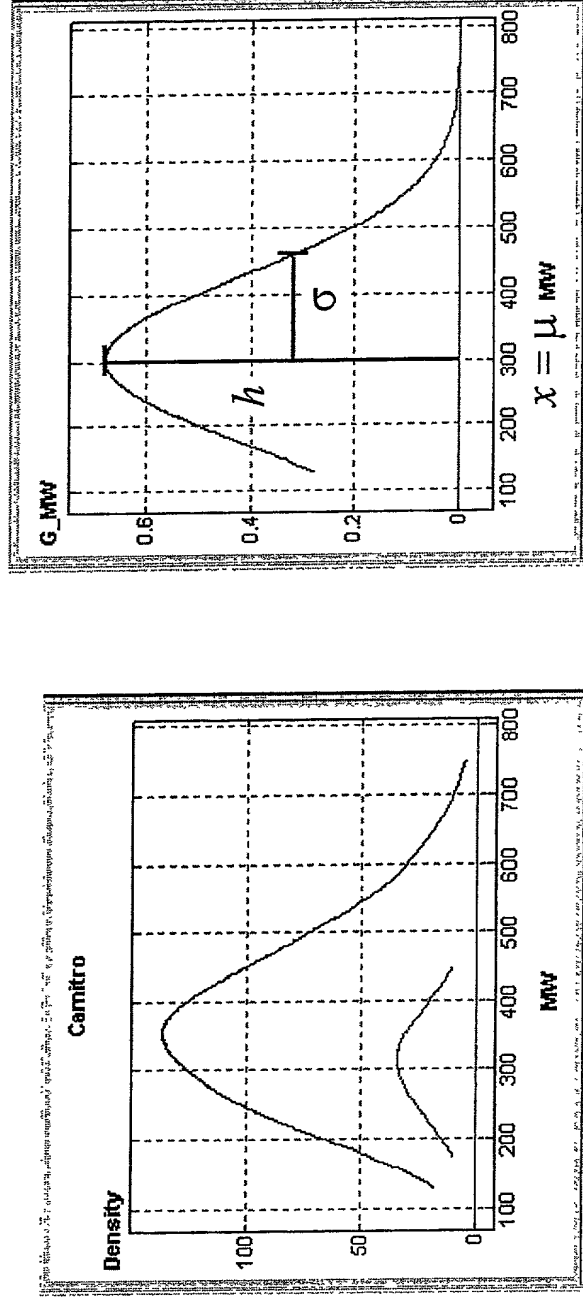


- - data point
- x - smoothed point
- - Smoothed data fit to a transformation function

Legend

Figure 3

Optimum Molecular Weight



$$g(x) = h e^{-\frac{(x-\mu)^2}{4\sigma^2}}$$

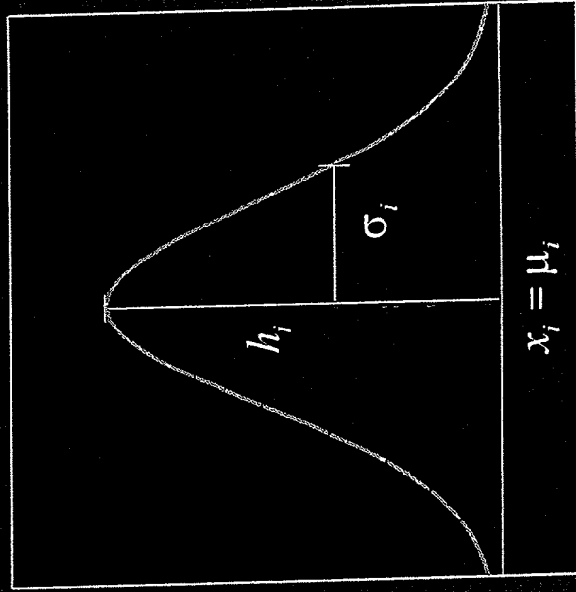
Figure 4A

Camitro

N-Dimensional Gaussian Modeling

Additive

$$g(x_{i..N}) = \frac{1}{N} \sum_{i=1}^N h_i e^{-\frac{(x_i - \mu_i)^2}{4\sigma_i^2}}$$



Multiplicative

$$g(x_{i..N}) = h e^{-\frac{1}{N} \sum_{i=1}^N (x_i - \mu_i)^2 / 4\sigma_i^2}$$

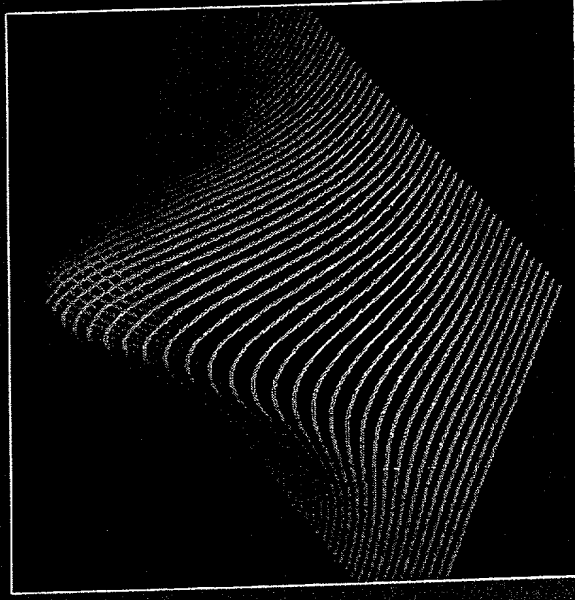


Figure 4 B

Camitro

Optimization Function

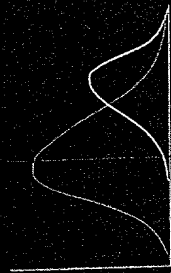
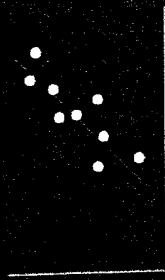
$$g(\mathbf{X}, \mu, \sigma, h, t) = t + h e^{-\sum_{k=1}^{N_x} (x_k - \mu_k)^2 / 4\sigma_k^2}$$

$$\begin{aligned} f = & S_{inh} \left[\frac{1}{N_{inh}} \sum_{i=1}^{N_{inh}} (g(\mathbf{X}_i, \mu, \sigma, h, t) - y_i)^2 \right] \\ & + S_{drug} \left[\frac{1}{N_{drug}} \sum_{j=1}^{N_{drug}} (g(\mathbf{X}_j, \mu, \sigma, h, t) - \bar{y}_{drug})^2 \right] \\ & + S_{fit} \left[\sigma_y^2 \sum_{k=1}^{N_x} \left(\frac{\mu_k - \mu_{0,k}}{range(\mathbf{X}_k^T)} \right)^2 + (t - t_0)^2 \right] \end{aligned}$$

Mean of the Squared Errors
of Inhibitor Affinity

Squared Error of the Means
of Drug Affinity

Constraints to prevent
Overfitting



Camitro

Figure 4C

Initial Values for Optimization

$$t_0 = \min(y)$$

$$h_0 = \max(y) - t_0$$

$$\mu_{0,k} = \frac{\sum_{i=1}^{N_{inh}} (y_i - t_0)^2 x_{k,i}}{\sum_{i=1}^{N_{inh}} (y_i - t_0)^2}$$

$$\sigma_{0,k} = \sqrt{\frac{\sum_{i=1}^{N_{inh}} (y_i - t_0)^2 (x_{k,i} - \mu_k)^2}{\sum_{i=1}^{N_{inh}} (y_i - t_0)^2}}$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{N_{inh}} (y_i - \bar{y}_{inh})^2}{N_{inh} - 1}}$$

Camitro

Figure 4D

Gaussian Optimization Function

$$f(\mathbf{x}) = t + h e^{-\sum_{k=1}^{N_x} (x_k - c_k)^2 / 4w_k^2}$$

$$f_o = S_y \left(\frac{\sum_{i=1}^{N_x} u_i (f(\mathbf{x}_i) - y_i)^2}{\sigma_y^2 \sum_{i=1}^{N_x} u_i} \right)$$

$$+ S_c \sum_{k=1}^{N_x} \frac{(c_k - c_{0,k})^2}{\sigma_{x_k}^2}$$

$$+ S_w \sum_{k=1}^{N_x} \frac{\sigma_{x_k}}{w_k}$$

$$+ S_t \frac{(t - t_0)^2}{\sigma_y^2}$$

Weighted Mean Squared Error

Center Constraint

Width (Focus) Constraint

Tare Constraint

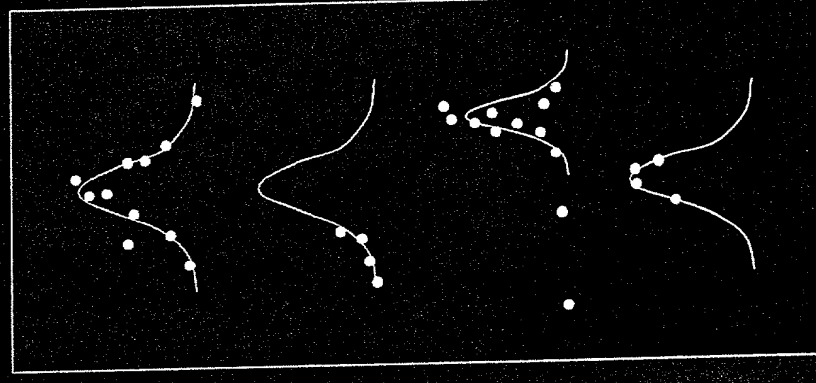


Figure 4E

Gaussian Optimization Starting Values

$$\sigma_y^2 = \frac{\sum_{i=1}^{N_{inh}} u_i (y_i - \bar{y})^2}{\sum_{i=1}^{N_{inh}} u_i}$$

$$t_0 = \min(y)$$

$$h_0 = \max(y) - t_0$$

$$\sigma_{x_k}^2 = \frac{\sum_{i=1}^{N_{inh}} u_i (x_{k,i} - \bar{x}_k)^2}{\sum_{i=1}^{N_{inh}} u_i}$$

$$v_i = \frac{(y_i - t_0)^2}{\sigma_y^2}$$

$$c_{0,k} = \frac{\sum_{i=1}^{N_{inh}} u_i v_i x_{k,i}}{\sum_{i=1}^{N_{inh}} u_i v_i}$$

$$w_{0,k}^2 = \frac{\sum_{i=1}^{N_{inh}} u_i v_i (x_{k,i} - c_{0,k})^2}{\sum_{i=1}^{N_{inh}} u_i v_i}$$

Camitro

Figure 4F

Performance Metrics

$$n_k = \frac{\sigma_{x_k}}{w_k}$$

Descriptor Focus

$$s = \sqrt{\frac{\sum_{i=1}^{N_r} u_i (f(x_i) - y_i)^2}{\sum_{i=1}^{N_r} u_i}}$$

Standard Error

$$r^2 = \frac{\left(\sum_{i=1}^{N_r} u_i (f(x_i) - \bar{f}(x)) (y_i - \bar{y}) \right)^2}{\sum_{i=1}^{N_r} u_i (f(x_i) - \bar{f}(x))^2 \sum_{i=1}^{N_r} u_i (y_i - \bar{y})^2}$$

Correlation Coefficient

$$q^2 = 1 - s^2 / \sigma_y^2$$

Residual Error

Camitro

Figure 46

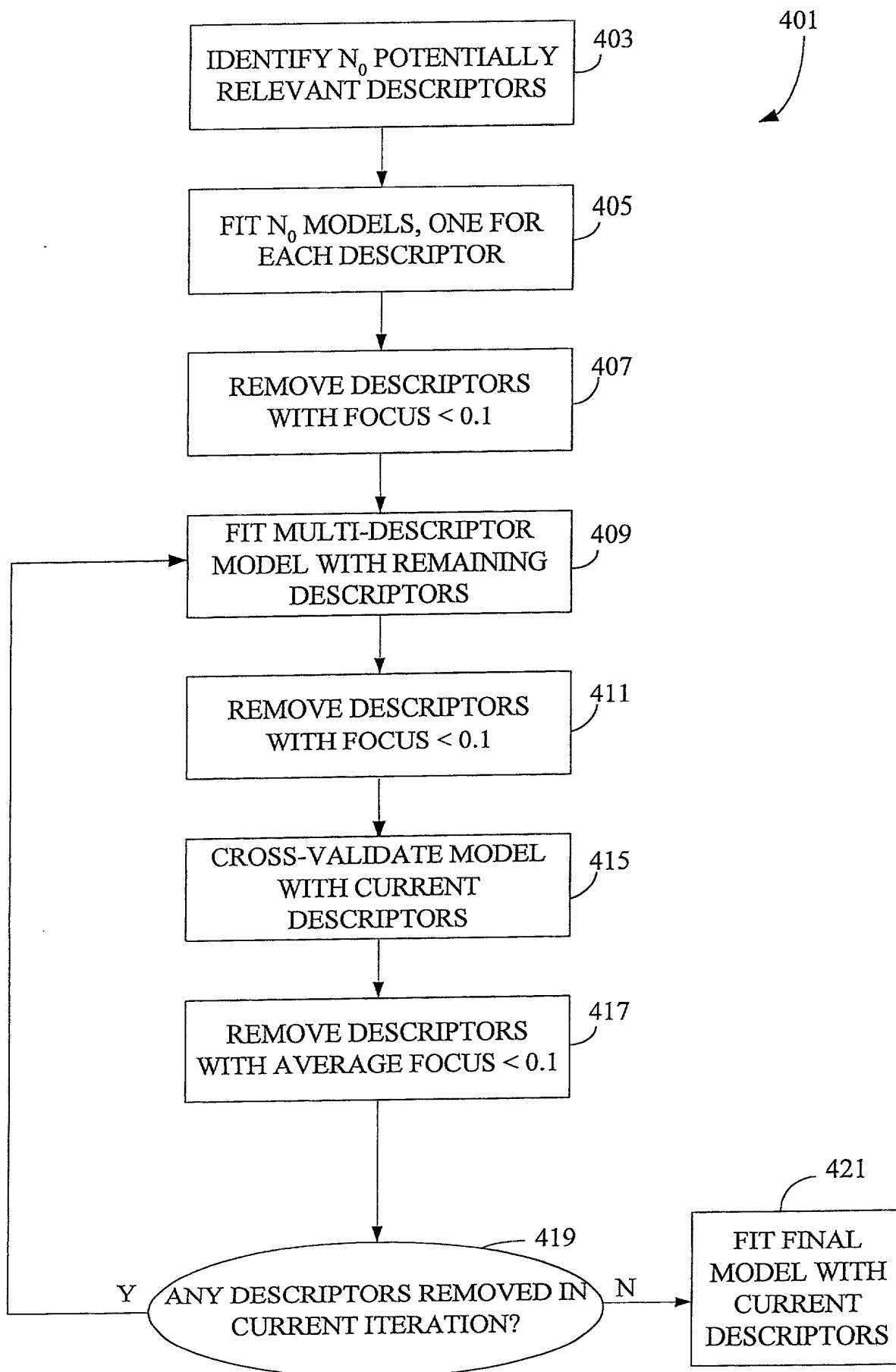


Figure 4H

Sigmoid Optimization Function

$$f(\mathbf{x}) = t + \frac{h}{1 + \sum_{k=1}^{N_x} e^{-n_k(x_k - c_k)}}$$

$$f_o = s_y \left(\frac{\sum_{i=1}^{N_y} u_i (f(\mathbf{x}_i) - y_i)^2}{\sigma_y^2 \sum_{i=1}^{N_y} u_i} \right)$$

Weighted Mean Squared Error

$$+ s_c \sum_{k=1}^{N_x} \frac{(c_k - c_{0,k})^2}{\sigma_{x_k}^2}$$

Center Constraint

$$+ s_n \sum_{k=1}^{N_x} |n_k| \sigma_{x_k}$$

Focus Constraint

$$+ s_t \frac{(t - t_0)^2}{\sigma_y^2}$$

Tare Constraint

Camitro

Figure 4I

Sigmoid Optimization Starting Values

$$t_0 = \min(y)$$

$$h_0 = \max(y) - t_0$$

$$v_i = \frac{(y_i - t_0)^2}{\sigma_y^2}$$

$$v'_i = \frac{(h_0 + t_0 - y_i)^2}{\sigma_y^2}$$

$$c_{h,k} = \frac{\sum_{i=1}^{N_{inh}} u_i v_i x_{k,i}}{\sum_{i=1}^{N_{inh}} u_i v_i}$$

$$c_{l,k} = \frac{\sum_{i=1}^{N_{inh}} u_i v'_i x_{k,i}}{\sum_{i=1}^{N_{inh}} u_i v'_i}$$

$$c_{0,k} = \frac{c_{h,k} + c_{l,k}}{2}$$

$$w_{h,k} = \sqrt{\frac{\sum_{i=1}^{N_{inh}} u_i v_i (x_{k,i} - c_{h,k})^2}{\sum_{i=1}^{N_{inh}} u_i v_i}}$$

$$w_{l,k} = \sqrt{\frac{\sum_{i=1}^{N_{inh}} u_i v'_i (x_{k,i} - c_{l,k})^2}{\sum_{i=1}^{N_{inh}} u_i v'_i}}$$

$$n_{0,k} = \frac{c_{h,k} - c_{l,k}}{w_{h,k} w_{l,k}}$$

Camitro

Figure 4J

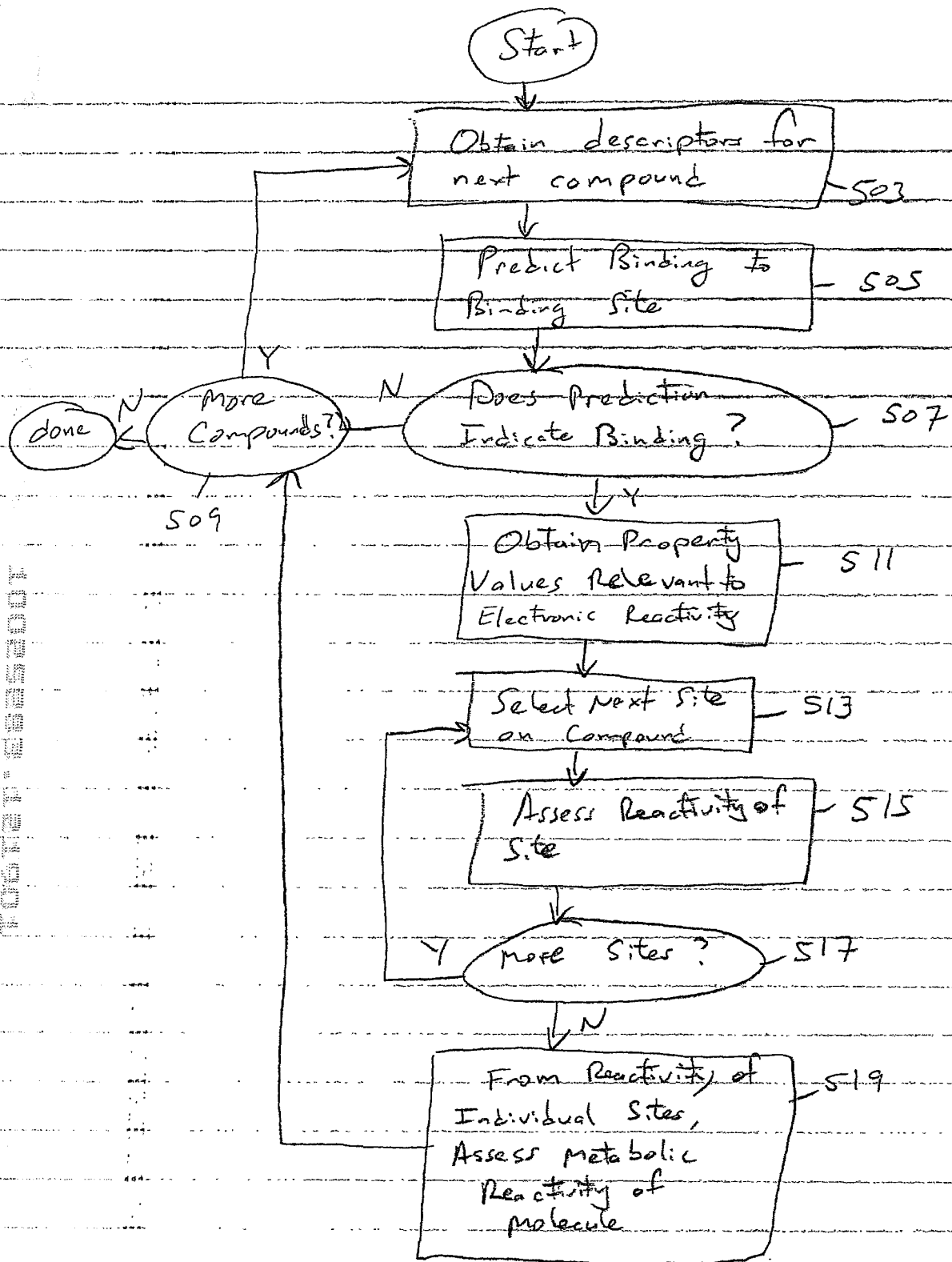


Figure 5

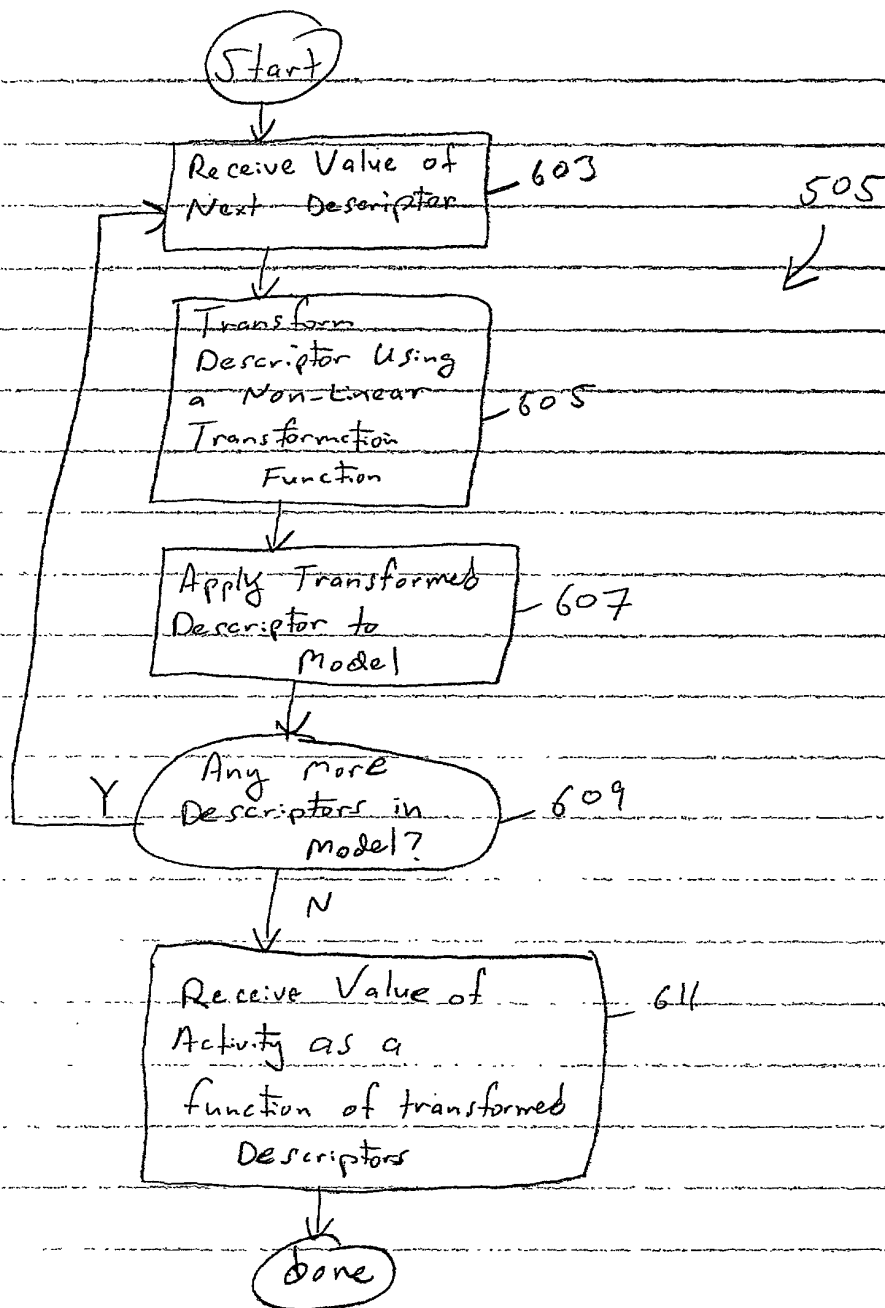
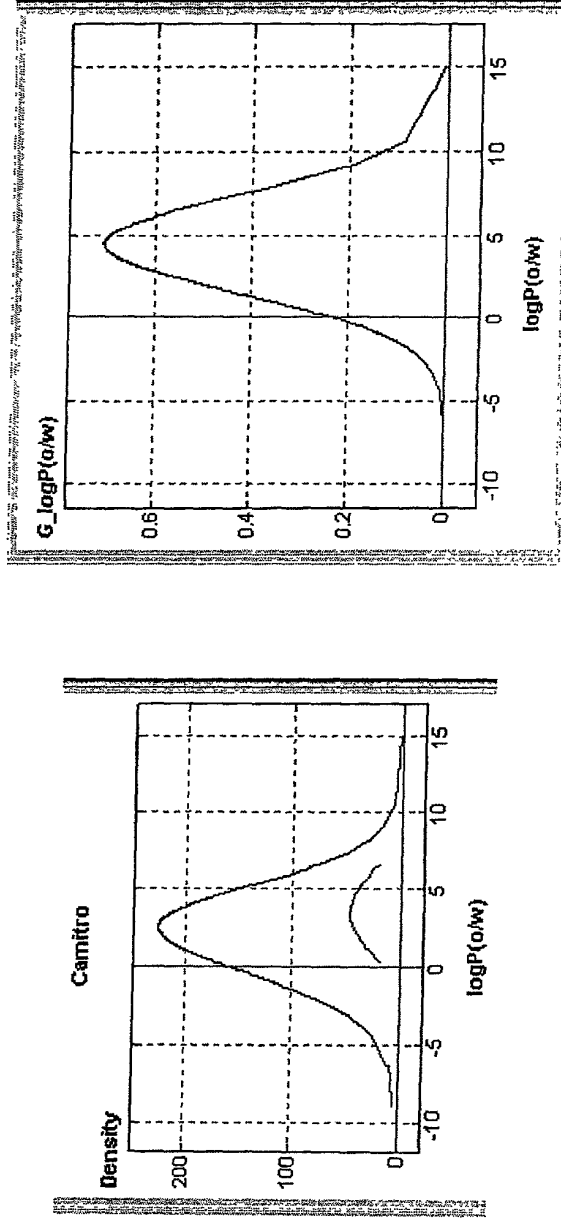


Figure 6

Optimum logP



Camitro

Figure 7A

Optimum Formal Charge

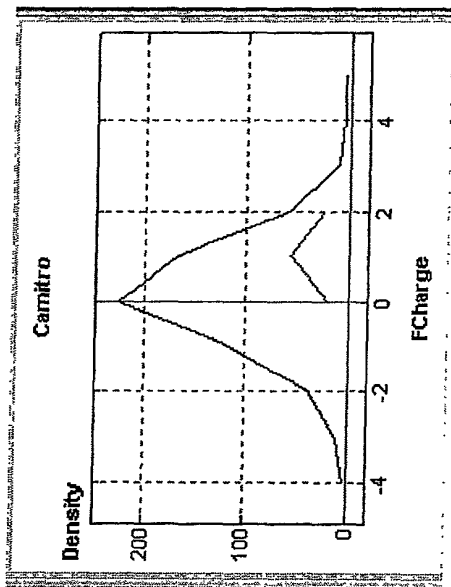
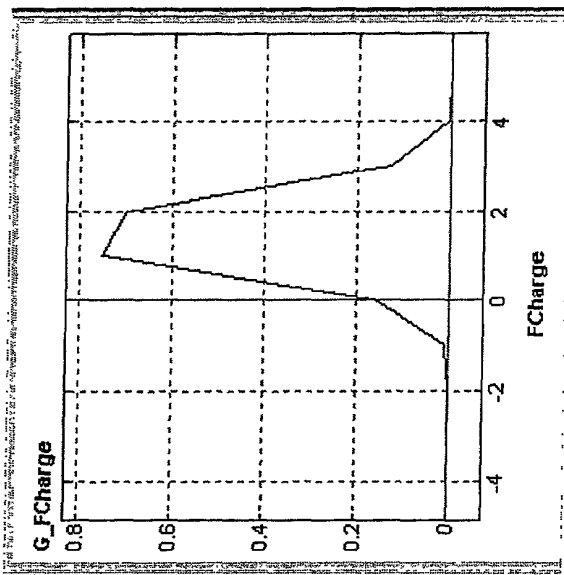


Figure 7B

Camitro

Automated Gaussian Fit

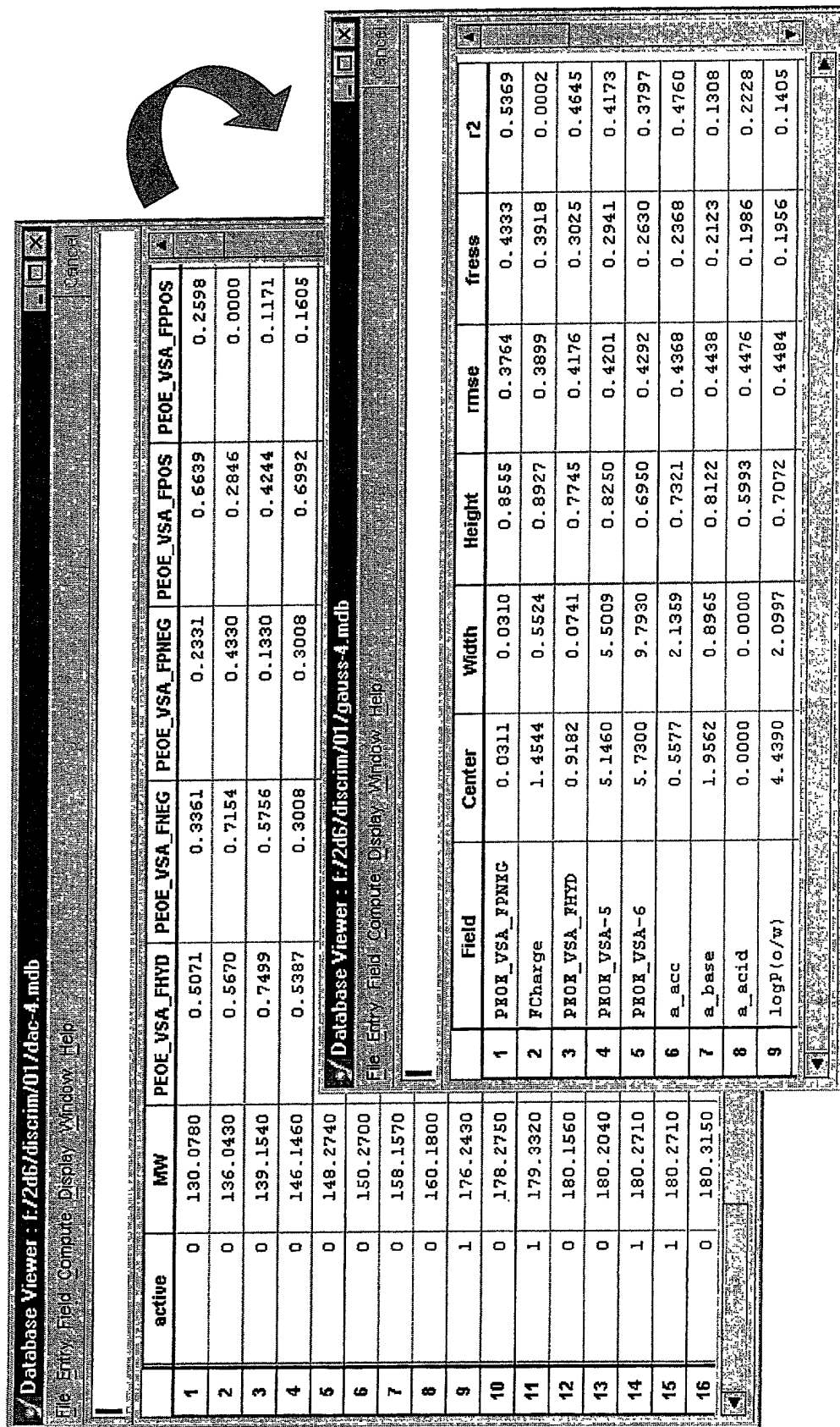


Figure 7C

2D6 K_i Model

Non-linear Size Relation

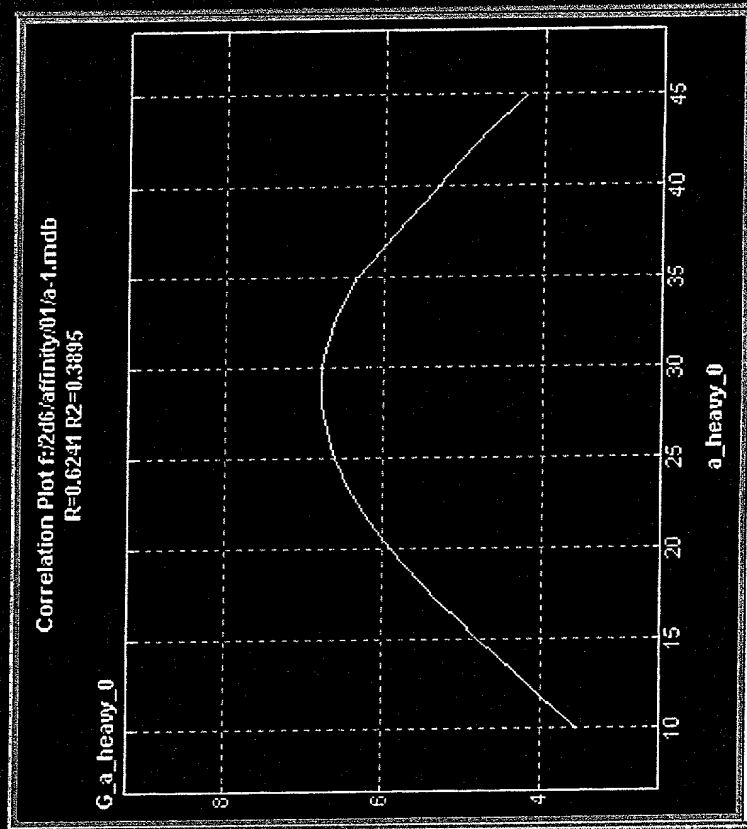
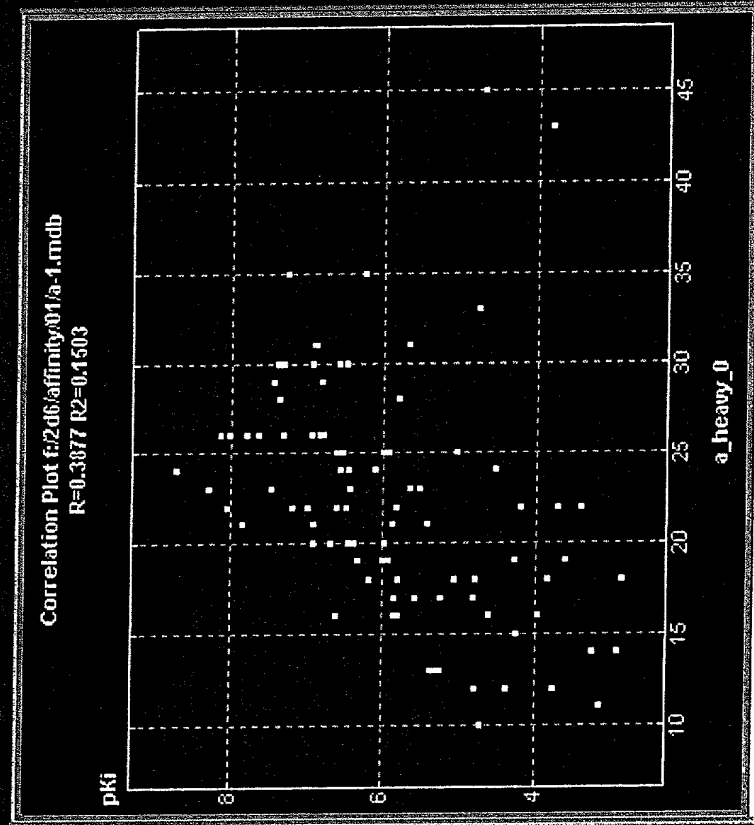


Figure 7D

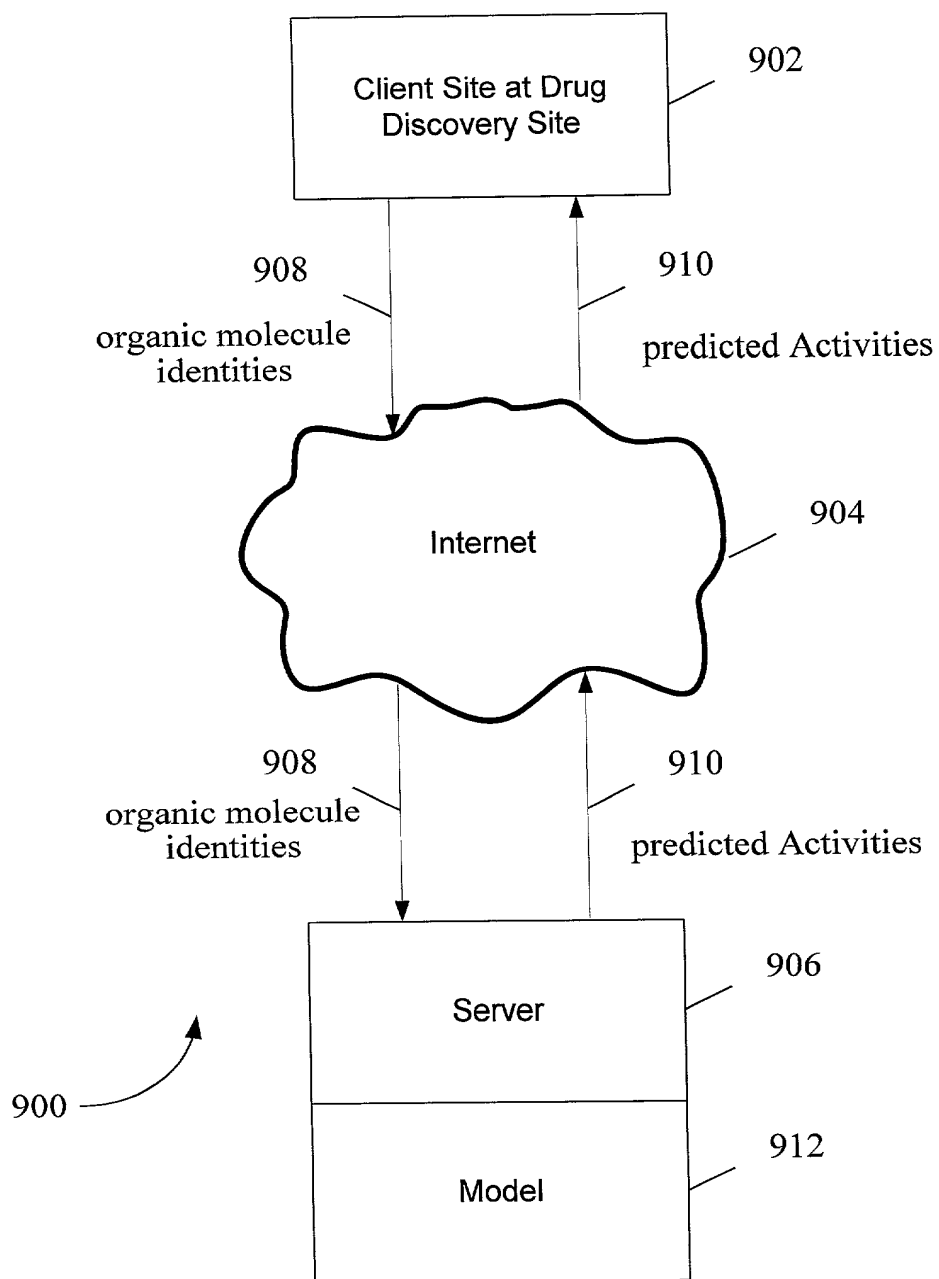


FIGURE 9

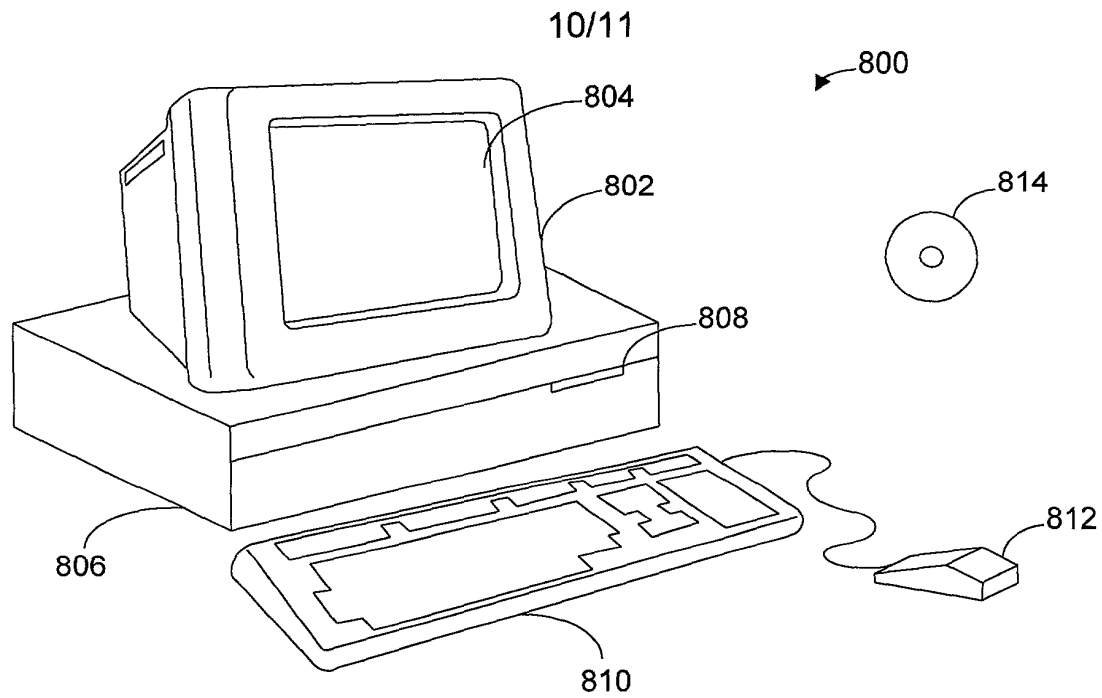


Figure 8A

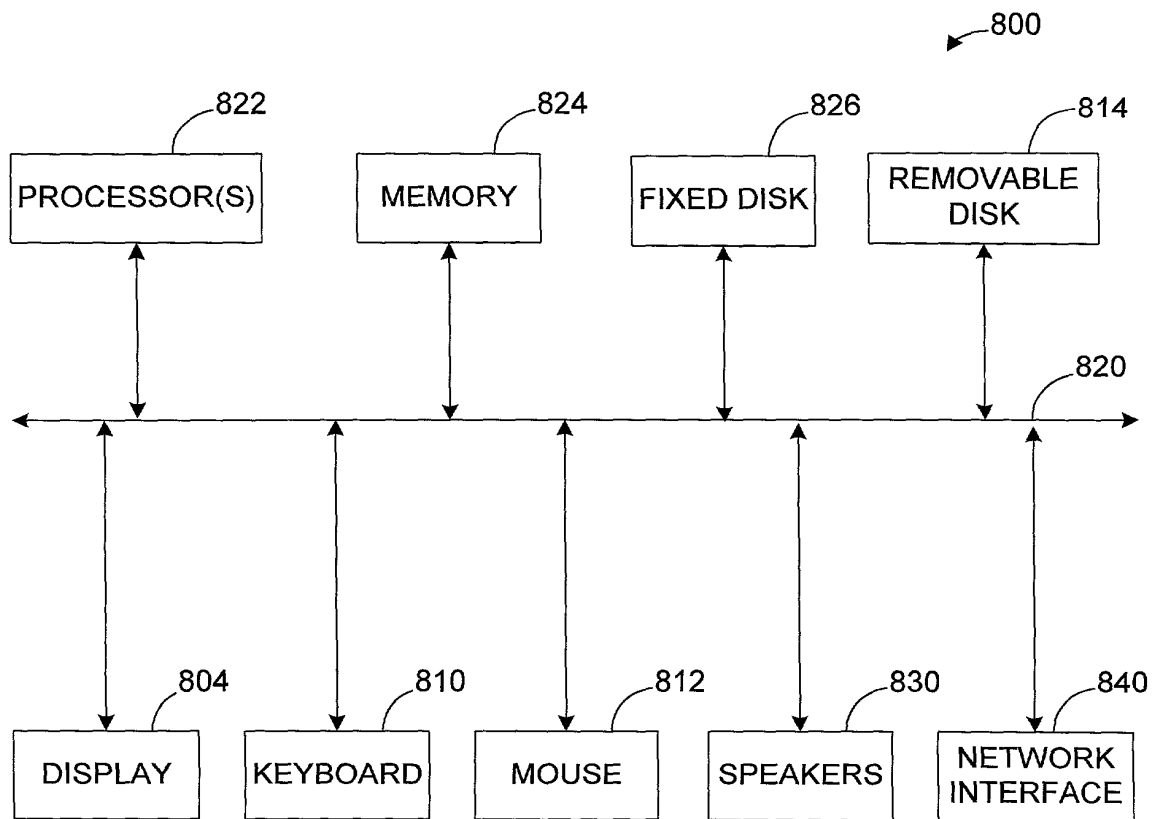


Figure 8B